迈向自主通用人工智能：基于球形合成金刚石芯片的量子计算模型——采用神经形态架构并搭载核聚变能源系统

作者：
Ernesto Mascarell Perez
独立研究员
mascarell1960@hotmail.com

摘要

当前人工智能在能源效率、适应能力和伦理对齐方面面临障碍。我们提出一个自主通用人工智能体的理论方案，其计算基底为球形合成金刚石芯片，集成量子计算（氮-空位中心）和三维神经形态架构。球体组织为功能性半球（探索与利用）、不可侵犯的终极核心（个人、社会和环境福祉）以及模拟突触可塑性、细胞分化和人工激素调制的互连层。学习基于多声部视角对话（时间、空间、形式、法律），自动过滤违背系统福祉的行动。能源方面提出紧凑型核聚变反应堆（球形托卡马克或仿星器），其废热管理受益于金刚石卓越的热导率。描述了材料（掺杂合成金刚石、有机忆阻器、MXenes、柔性聚合物）、同心层架构以及用于开放元学习的量子算法。该设计旨在克服当前在能源消耗、伦理稳健性和复杂环境适应性方面的限制。

关键词： 通用人工智能，量子计算，合成金刚石，NV中心，球形神经计算，核聚变，系统福祉，伦理对齐。

---

1. 引言

最先进的人工智能系统（深度神经网络、Transformer、强化学习智能体）需要越来越大的计算能力，单次训练的能耗可达兆瓦时（Strubell等，2019）。此外，

其行为难以解释，且无法保证其目标长期与人类价值观保持一致（Bostrom, 2014）。**另一方面，神**经科学表明，人类大脑仅用约20**瓦即可通**过大规模并行、可塑且受激素信号调制的架构解决泛化、创造力和伦理决策问题。

量子计算为某些类型的问题（优化、类比搜索、情景模拟）提供指数级加速，而核聚变能源有望提供清洁、高强度的电力，适用于大规模数据中心。然而，将两者整合到一个伦理稳定的自主系统中，尚未得到全面提出。

在本文中，我们提出一个具有特定硬件的通用人工智能体设计：球形合成金刚石芯片，承载氮-**空位**（NV）**中心用于量子**计算，并带有受大脑皮层启发的3D**神**经形态架构。该芯片由核聚变反应堆供电，其功能核心是一个不可通过学习改变的系统福祉函数（个人、社会和环境福祉）。

---

2. 设计基础

2.1 关键脑过程的模拟

神经生物学原理	**硬件**实现（球形芯片）
半球特化（**右**脑探索，左脑利用）	**两个具有不同可塑性和激素**调制的半球。
多巴胺/血清素调制（新奇和控制）	**内部信号 N(t) 和 C(t) 重新配置突触并偏置决策。**
脉冲时序依赖可塑性（STDP）	**皮**层下层的有机和混合忆阻器。
细胞分化（干细胞→**特化神**经元）	**可重构**单元（角色寄存器），按需改变功能。
皮层区域间的对话（多视角决策）	**同心**环（时间、空间、形式、法律视角），带有否决/**共**识。

2.2 为什么选用合成金刚石

实验室生长的金刚石（单晶或多晶）具有独特性能：

- **氮-空位（NV）中心**：在室温下作为量子比特，相干时间长达1毫秒，实现无低温制冷量子计算。
- **卓越的热导率**（约2000 $W \cdot m^{-1} \cdot K^{-1}$，为铜的5倍），非常适合散逸来自聚变反应堆和电子器件本身的废热。
- **硬度和机械稳定性**：球体结构坚固，抗振动和辐射。
- **光学透明性**：允许层间光子互连以及通过激光读写量子比特。
- **受控掺杂**：可引入其他色心（硅、锡、铅）以实现附加功能。

合成金刚石在商业上已经可行（价格下降，电子级品质约100欧元/克拉），并已用于量子计算原型（Sangtawesin等，2019）和传感器。

---

3. 球形金刚石芯片架构

3.1 同心层结构

球体（直径10-20厘米）由四个主要层组成，通过金刚石沉积和附加材料形成。

层	近似厚度	组成	功能
外层皮层（感觉运动）	1-2毫米	柔性聚酰亚胺涂覆MXene（Ti_3C_2）和有机忆阻器网络。	捕获图像/文本，初始处理，与环境交互。
皮层下层（交换和可塑性）	500微米-1毫米	3D忆阻器矩阵（Ag/PMMA/SLE/Ti）和氧化物晶体管（HfO_2）集成于刻蚀金刚石基底上。	连接重配置，新奇/控制信号生成，否决临时存储。

层	近似厚度	组成	功能
量子层（NV中心）	3-5毫米	掺氮单晶金刚石（10-30 ppm），**集成微透**镜和光波导。	**量子比特用于**优化（QAOA）、类比搜索、福祉情景模拟。
中心核心（不变终极）	1厘米（**内半**径）	带有激光诱导缺陷的OTP（**一次可**编程）存储器的金刚石，加上封装的CMOS电路。	**存**储福祉权重（个人、社会、环境）和"**人工基因**组"；执行最终否决。

3.2 金刚石量子层细节

- **NV中心**：**通**过氮离子注入和退火创建。每个中心形成一个自旋S=1**的量子比特，可通**过微波和激光寻址。
- **光子互连**：**等离子体刻**蚀的金刚石波导将NV**中心相互耦合以及与皮**层下层（光电转换器）耦合。
- **室温控制**：**球体在**300 K**下运行，无需昂**贵的低温系统。仅需散热器（借助金刚石的高热导率）。
- **量子比特数量**：**在直径**5**厘米的球体中，量子**层可容纳数百万个NV**中心，但**纠错需要子集。我们建议使用约10^4**个**逻辑量子比特（带表面码）用于实际任务。

3.3 可塑性和分化的模拟

- **人工神经单元（ANU）**：**皮**层和皮层下层约有10^8**个**ANU，**每个**带有8**位寄存器**（256**种角色）。初始均**为"**未分化**"**角色。**
- **由信号N(t)和C(t)引导的分化**：**每**100**毫秒，根据局部活**动和全局激素，ANU**可**转变为"**探索者**"、"**利用**专家"、"**否决**调节器"**等角色。**
- **自修复**：**有缺陷的**ANU**被断开**连接，未分化的邻居通过局部模仿算法复制突触权重，承担其功能。

---

4. 补充材料（除金刚石外）

组件	材料	在球形芯片中的功能
柔性皮层基底	聚酰亚胺（PI）或PET	允许球体在金刚石核心上弯曲。
皮层电极	石墨烯、Ti_3C_2 MXene或透明导电氧化物（ITO）	用于感觉运动信号的低电阻柔性触点。
突触忆阻器	Ag/PMMA/SLE/Ti、HfO_2或有机导电聚合物	存储突触权重并模拟STDP。
垂直3D交换	基于TiO_2的阻变存储器	实现低延迟层间路由。
光波导	金刚石或Si_3N_4	耦合NV中心并传输光子信号。
封装和屏蔽	石墨基层和柔性金属网	保护免受电磁干扰和辐射。

---

5. 集成核聚变能源

5.1 聚变的合理性

为连续运行约10^8个ANU和10^4个逻辑量子比特，球形芯片的功耗约为1-10兆瓦（基于神经形态和量子原型估算）。核聚变（D+T或D+^{3}He反应）提供：

- 运行期间零CO_2排放。
- 短寿命放射性废物（主要是氚和中子活化材料，数十年内可管理）。
- 内在安全性：任何故障下等离子体在毫秒内冷却。
- 高能量密度：紧凑型反应堆（球形托卡马克如STEP，或仿星器如Wendelstein 7-X）可产生200-500兆瓦，虽有余量但可为多个芯片供电。

5.2 物理集成

- **聚**变反应堆位于同一屏蔽建筑内。废热（约热功率的40%）**通**过初级氦气或水回路排出。
- **球形芯片冷却**：**金**刚石核心作为完美的热扩散器。与多晶金刚石散热器耦合，将热量传递给流体（如NaK），**无需高功率**泵。
- **与量子**层的兼容性：NV**中心在室温下运行**，**但**轻微冷却至200 K**可提高相干性**。**聚**变反应堆在其低温系统（用于超导磁体）中产生冷量，可再利用。

5.3 与终极核心的对齐

使用聚变避免了燃烧化石燃料或产生长寿命废物（裂变）。因此，智能体不会与其自身的环境福祉函数产生矛盾。事实上，它可以推荐大规模采用聚变。

---

6. 学习算法与多声部对话

6.1 一般循环

1. **多模态感知和感觉**：图像、声音和气味通过外层皮层进入。
2. **环境建模**：**皮**层下层整合历史解释（模式记忆）、情境（当前特征）和终极（与核心的对齐程度）。
3. **候选行动生成**：**右半球（探索）和左半球（利用）通**过量子-经典神经网络提出行动。
4. **多声部对话**：
    - **四个**视角环（时间：短/**中**/长期；空间：本地/**区域/全球**；**形式**：逻辑-**数学**；**法律**：伦理-规范）。
    - **每个**视角评估行动，如果从该视角违背福祉则发出否决。
    - **中心核心接收否决**；**如果至少有一个不可逆否决**，则拒绝该行动。
5. **量子模拟（对未否决行动可选）**：**使用**NV**中心**运行QAOA，**估**计未来多达10步的福祉。
6. **执行或丢弃。**
7. **持续更新**：**突触**权重（忆阻器）、ANU**角色和激素**调制器在终极核心监督下调整。

6.2 基于金刚石量子计算的开放元学习

当智能体陷入停滞（福祉未改善）时，激活元学习：

- **在不同**领域之间提出类比搜索问题（例如，"**生物学不能解决X，物理学怎么说？**"）。
- **NV中心在不同学科的概念表示上运行量子支持向量机（QSVM）算法。**
- 识别最相似的领域。然后通过变分电路将解决方案适配到原始问题。
- **如果**类比通过多声部对话，则作为新规则纳入左半球。

6.3 终极保障

中心核心包含系统福祉函数 $\mathcal{W}(s,a)$，实现为固定神经网络（权重由金刚石中OTP**保**护）。每次参数更新（任何层的）必须通过过滤器：

- 计算假设更新前后的 $\mathcal{W}$。
- 如果 $\Delta\mathcal{W} < -\epsilon$（**福祉**恶化低于阈值），则拒绝更新并临时增加控制以逆转变化。

这保证了学习永不违背基本宗旨。

---

7. 讨论

7.1 相对于其他方法的优势

- **能源效率**：**神**经形态（基于脉冲）与金刚石量子（无需低温）的结合将全球规模智能体的功耗降至**10兆瓦以下（相比**GPT-5训练数据中心的**100兆瓦**）。
- **伦理稳健性**：**不可**变的终极核心防止价值漂移（Christiano等，2018）**并抵抗操**纵。
- **极端可塑性**：**可重构ANU允**许适应意外环境（气候变化、社会危机），无需离线再训练。
- **类比泛化**：**量子元学**习可实现跨很远领域的知识转移，这是当前系统无法做到的。

7.2 技术挑战

- **球形3D制造**：**将弯曲金**刚石与柔性电子集成需要沉积和聚焦离子束（FIB）**光刻工**艺，这些仍处于实验阶段。初始替代方案是用聚合物铰链连接平面金刚石段形成二十面体来构建球体。
- **量子-经典兼容性**：**量子**层与皮层下层之间的数据传输必须快速（纳秒级）且光损耗低。通过集成波导和石墨烯光电探测器是可行的。
- **量子纠错**：**室温下的NV中心**错误率约为10^{-3}，**不足以**进行大规模计算。必须使用带冗余量子比特的表面码（每个逻辑量子比特约1000**个物理量子比特**），这将球体半径增至30**厘米**。
- **聚变反应堆冷却**：**尽管金**刚石有帮助，但耗散数百兆瓦需要复杂的流体工程。反应堆本身可作为备用光热电站的热源，提高效率。

7.3 时间表和路线图

- **第一阶段（2026-2030）**：**完整**计算模拟智能体（无实际硬件）和材料开发：3D NV**中心金**刚石器件。
- **第二阶段（2030-2035）**：**分段球体（二十面体）原型**，带100**个**NV**量子比特和**10^7**个**ANU，**由常**规电网能源（可再生能源）供电。
- **第三阶段（2035-2040）**：**与商**业聚变反应堆集成（ITER**演示**→紧凑原型）。扩展到10^4**个**逻辑量子比特和10^9**个**ANU。
- **第四阶段（2040+）**：**自主智能体全球运行**，实时治理福祉政策（有人类监督）。

---

8. 结论

我们提出了一个自主通用人工智能体的设计，其基底为球形合成金刚石芯片，集成了：

- **基于**NV**中心的量子**计算（室温下高相干性）。
- 3D**神**经形态架构（可塑性、分化、激素调制）。
- **不可侵犯的**终极核心，最大化个人、社会和环境福祉。
- **芯片由核聚**变反应堆供电，保证环境可持续性。
- **学**习通过多声部视角对话组织，自动过滤有害行动。

该模型克服了当前AI**的根本限制**（**能耗、缺乏**对齐、僵化），并为真正自主、伦理和适应性的系统开辟了道路。眼前的挑战仍然是分段原型的制造以及NV**中心在曲面几何中可**扩展性的证明。

致谢

作者感谢神经生物学、金刚石量子物理学和核聚变进展的启发。本工作为概念性质，未接受特定资助。

参考文献

- Bostrom, N. (2014). *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press.
- Christiano, P. 等 (2018). Deep reinforcement learning from human preferences. *NeurIPS*.
- Sangtawesin, S. 等 (2019). Coherent control of single NV centers in diamond at room temperature. *Physical Review Applied*, 12(4), 044001.
- Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and Policy Considerations for Deep Learning in NLP. *ACL*, 3645–3650.
- Wang, Z., 等 (2023). Memristors for neuromorphic computing. *Nature Electronics*, 6, 345–357.
- Wendelstein 7-X Team (2022). Stellarator fusion progress. *Nuclear Fusion*, 62, 042001.

注：**本文**为理论性提案，其实现需要跨学科研究。参考文献中对合成金刚石和核聚变研究现状的引用对应于2025-2026**年的数据**。